The Canadian Journal of Statistics
Vol. 29, No. ?, 2001, Pages ???-???
La revue canadienne de statistique

Statistical analyses for round robin interaction data

Paramjit S. GILL and Tim B. SWARTZ

Key words and phrases: Bayesian methods; components of covariance; linear mixed models; Markov chain Monte Carlo; parametric bootstrap; social relations models.
AMS 1991 subject classifications: Primary 62J05; secondary 62F15, 62P25.

ABSTRACT

This paper considers the analysis of round robin interaction data whereby individuals from a group of subjects interact with one another, producing a pair of outcomes, one for each individual. The authors provide an overview of the various analyses applied these types of data and extend the work in several directions. In particular, they provide a fully Bayesian analysis for such data, and use a real data example for illustration purposes.

RÉSUMÉ

Cet article concerne l'analyse de données issues de tournois à la ronde dans lesquels l'interaction des membres d'un groupe deux à la fois donne lieu à des paires de résultats, un pour chaque individu. Les auteurs passent en revue les différents types d'analyses existantes pour ce type de données et en proposent diverses généralisations. Ils montrent notamment comment analyser de telles données par une approche bayésienne. Leur propos est illustré par un exemple concret.

1. INTRODUCTION

Round robin interaction data arise when individuals from a group of subjects interact with one-another. An interaction between two subjects produces a pair of outcomes, one for each subject. We assume that these outcomes, measured on a numerical scale, represent some phenomenon under study. For example, social psychologists study perceptions and interpersonal attractions amongst people. Each subject rates the others with respect to a set of social traits to give rise to a summary score (say between 0 and 100). In studies on the social behaviour of animals, a researcher may record the number of times an animal extends an activity (for example, grooming) towards another animal. The researcher may be interested in studying the differences between the social behaviour of male and female animals towards the same/different sex animals. In studies involving networking among large organizations (say business firms), one may count the number of executives from organization A serving on the board of directors of organization B and vice-versa. To study the intellectual influence of a professional journal A on another journal B, the number of citations by authors in journal B of work published in journal A is counted (Stigler 1994). Clearly, there are many types of scientific investigations that give rise to round robin interaction data, and as such, there is a need for good methods of statistical analysis.

In round robin experiments, each subject in the group plays the dual role of an "actor" and a "partner" giving rise to bivariate data. This renders a round robin experiment different than a typical two-factor experiment. Another important characteristic of round robin designs is that a subject does not interact with oneself. Thus, the diagonal cells in the data matrix are missing by design. Although the first formal statistical model for round robin interaction data was introduced by Lev & Kinder (1957), a more general and the most commonly cited model was studied by Warner, Kenny & Stoto (1979). Further developments of the model as a social relations model are discussed in a series of papers and a book by David Kenny (Kenny 1994). Related "network models" are discussed by Wasserman & Faust (1994).

Suppose that a round robin design involves m subjects where subjects i and j meet n_{ij} times, $i \neq j$. If some n_{ij} are zero, the design is said to be incomplete. If all n_{ij} are equal, we say that the design is balanced. For unequal n_{ij}, the design is unbalanced.

When subjects i and j meet on the kth occasion, we obtain a pair of observations y_{ijk} and y_{jik} as a realization of continuous random variables. Here y_{ijk} represents the response of subject i as an *actor* towards subject j as a *partner* on the kth occasion; and in y_{jik}, the roles are reversed. Warner *et al.* (1979) proposed the general round robin model

$$\begin{aligned} y_{ijk} &= \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \\ y_{jik} &= \mu + \alpha_j + \beta_i + \gamma_{ji} + \varepsilon_{jik} \\ k &= 1, \ldots, n_{ij};\ 1 \leq i \neq j \leq m. \end{aligned} \tag{1}$$

In this model μ is the general mean; α_i represents the effect of subject i as an actor; β_j is a partner effect due to subject j; γ_{ij} is an interaction effect representing the special adjustment which subject i makes for subject j; ε_{ijk} represents the error term which picks up the measurement error and/or variability in behaviour at different occasions. Note that if $n_{ij} < 2$ for some i, j, the data does not distinguish between γ_{ij} and ε_{ijk}. Except for the general mean μ, all parameters in model (1) are assumed to be random variables and are known as the random effects. It is assumed that

$$\begin{aligned} \mathrm{E}(\alpha_i) &= \mathrm{E}(\beta_j) = \mathrm{E}(\gamma_{ij}) = \mathrm{E}(\varepsilon_{ijk}) = 0 \\ \mathrm{var}(\alpha_i) &= \sigma^2_\alpha,\ \mathrm{var}(\beta_j) = \sigma^2_\beta,\ \mathrm{var}(\gamma_{ij}) = \sigma^2_\gamma,\ \mathrm{var}(\varepsilon_{ijk}) = \sigma^2_\varepsilon \\ \mathrm{cov}(\alpha_i,\ \beta_i) &= \sigma_{\alpha\beta},\ \mathrm{cov}(\gamma_{ij}, \gamma_{ji}) = \sigma_{\gamma\gamma},\ \mathrm{cov}(\varepsilon_{ijk}, \varepsilon_{jik}) = \sigma_{\varepsilon\varepsilon} \end{aligned} \tag{2}$$

and all other covariances are assumed equal to zero. We call the set of parameters $\{\sigma^2_\alpha, \sigma^2_\beta, \sigma^2_\gamma, \sigma^2_\varepsilon, \sigma_{\alpha\beta}, \sigma_{\gamma\gamma}, \sigma_{\varepsilon\varepsilon}\}$ the variance-covariance parameters (or components) of the round robin model and these are the primary parameters of interest. Instead of the covariance parameters $\{\sigma_{\alpha\beta}, \sigma_{\gamma\gamma}, \sigma_{\varepsilon\varepsilon}\}$, one may alternatively use the correlation coefficients $\{\rho_1, \rho_2, \rho_3\}$ as defined by $\sigma_{\alpha\beta} = \rho_1\sigma_\alpha\sigma_\beta$, $\sigma_{\gamma\gamma} = \rho_2\sigma^2_\gamma$ and $\sigma_{\varepsilon\varepsilon} = \rho_3\sigma^2_\varepsilon$.

In some cases, it may be desirable to add parameters to the fixed effects part of the model. For example, when modelling sports data, a home-field advantage parameter may be required. The model as such, is appropriate for a single group of subjects. When we have more than one type of group (say, males and females), some of the model parameters may be group specific.

Warner *et al.* (1979) introduced ANOVA based estimation for the variance-covariance components in the case of balanced designs ($n_{ij} = n > 1$). In addition to point estimates, one may also wish to estimate standard errors. Bond & Lashley (1996) provide explicit formulas for standard errors for a subclass of the balanced round robin model (without the error component ε_{ijk}). In the unbalanced design case, ANOVA estimation becomes complex and it is practically impossible to derive standard errors for the ANOVA estimators. To have an appreciation of the degree of difficulty, the reader is referred to Searle, Casella & McCulloch (1992, Chapter 5).

Some notable properties of ANOVA estimation are unbiasedness and the simplicity of computations in the case of balanced designs. Also in the case of balanced designs, ANOVA estimators are known to have minimum variance among unbiased estimators. But many of these properties do not hold for the general design. It is well known that ANOVA can yield negative estimates of positive parameters. On the other hand, maximum likelihood estimators are known to have better large-sample properties such as large-sample normality and efficiency. With the advent of fast computational tools, the need for simplicity of computations should not be an impediment to using likelihood based methods.

Wong (1982) proposed maximum likelihood estimation (MLE) under the assumption that the random effects are normally distributed. He used the EM algorithm for maximum likelihood estimation in the case of balanced designs but remarked that the unbalanced case is more complicated. At the end of his paper, Wong remarked that, "a challenging piece of research... is the development of satisfactory and computationally feasible numerical methods for a... fully Bayesian analysis..." Recent advances in Monte Carlo Markov chain (MCMC) methods have empowered us to meet this challenge.

In Section 2, we review classical ANOVA based estimation and extend this method to unbalanced round robin designs. For estimating the standard errors of ANOVA estimators, we implement bootstrap methods. Although ultimately we advocate the use of Bayesian methods, ANOVA estimation is attractive because of the simplicity of computations. Also, ANOVA estimates may provide starting values for more complex approaches such as maximum likelihood estimation and Bayesian analyses via MCMC.

Section 3 contains discussion of maximum likelihood estimation and empirical Bayes estimation of random effects. The MLE algorithm proposed is based on the method of scoring. The method is straightforward to describe and implement. We note that the approach described in Section 3 directly maximizes the likelihood function whereas the approach of Wong (1982) produces restricted maximum likelihood (REML) estimates. An added advantage of the method of scoring over Wong's implementation of the EM algorithm is that it produces an estimate of the asymptotic covariance matrix for the maximum likelihood estimates.

In Section 4, we present a fully Bayesian analysis for round robin interaction data. The priors are based on the structure of the experiment and hence are not empirical Bayes. The computational techniques are based on the theory of Markov chains where Metropolis steps are embedded within the Gibbs sampling algorithm.

In Section 5, we apply the three methods to a real dataset. We compare the results of the various methodologies and address issues of implementation.

A general, though not surprising theme that emerges from this paper is that more realistic and satisfying models require more sophisticated and modern computational techniques. With the availability of fast computers, it is now possible to use sophisticated methodology on a routine basis.

2. ANOVA ESTIMATION

ANOVA based estimation is the approach that is routinely used for the analysis of round robin interaction data. Although the examples found in the social sciences literature typically deal with balanced designs, unbalanced designs often arise in practice. For example, in experiments involving a large number of individuals, it may not be feasible to pair all the subjects. For the sake of completeness, we extend the ANOVA approach for estimating the variance-covariance components in the general (i.e., unbalanced) case. ANOVA estimation is essentially the method of moments estimation: we equate a set of "sums of squares" to their expected values and solve the set of linear equations for the variance-covariance parameters. Due care is needed in selecting the sums of squares so that the resulting system of linear equations is non-singular. Apart from this requirement, no optimality criterion is employed in selecting the particular sums of squares.

We use familiar matrix notation: I_k is the $k \times k$ identity matrix; A' denotes the transpose, $|A|$ the determinant and A^{-1} the inverse of a matrix A.

Using the standard ANOVA notation, we define

$$n_{i.} = \sum_{j \neq i=1}^{m} n_{ij}, \quad n_{.j} = \sum_{i \neq j=1}^{m} n_{ij}, \quad 2N = \sum_{i=1}^{m} \sum_{j \neq i=1}^{m} n_{ij}$$

$$\bar{y}_{i..} = \frac{1}{n_{i.}} \sum_{j \neq i=1}^{m} \sum_{k=1}^{n_{ij}} y_{ijk}, \quad \bar{y}_{.j.} = \frac{1}{n_{.j}} \sum_{i \neq j=1}^{m} \sum_{k=1}^{n_{ij}} y_{ijk}, \quad \bar{y}_{ij.} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} y_{ijk}, \quad \bar{y}_{...} = \frac{1}{2N} \sum_{i=1}^{m} \sum_{j \neq i=1}^{m} \sum_{k=1}^{n_{ij}} y_{ijk}.$$

Let the 7×1 column vector $\theta = (\sigma^2_\alpha, \sigma^2_\beta, \sigma^2_\gamma, \sigma^2_\varepsilon, \sigma_{\alpha\beta}, \sigma_{\gamma\gamma}, \sigma_{\varepsilon\varepsilon})'$ denote the vector of unknown parameters. Define further, the set of sums of squares

$$SSA = \sum_{i=1}^{m} n_{i.} \left(\bar{y}_{i..} - \bar{y}_{...} \right)^2, \quad SSB = \sum_{j=1}^{m} n_{.j} \left(\bar{y}_{.j.} - \bar{y}_{...} \right)^2,$$

$$SSAB = \sum_{i=1}^{m} n_{i.} \left(\bar{y}_{i..} \bar{y}_{...} \right) \left(\bar{y}_{.i.} - \bar{y}_{...} \right), \quad SSG = \sum_{i=1}^{m} \sum_{j \neq i=1}^{m} n_{ij} \left(\bar{y}_{ij.} - \bar{y}_{...} \right)^2,$$

$$SSGG = \sum_{i=1}^{m} \sum_{j \neq i=1}^{m} n_{ij} \left(\bar{y}_{ij.} - \bar{y}_{...} \right) \left(\bar{y}_{ji.} - \bar{y}_{...} \right), \quad SSE = \sum_{i=1}^{m} \sum_{j \neq i=1}^{m} \sum_{k=1}^{n_{ij}} \left(y_{ijk} - \bar{y}_{ij.} \right)^2,$$

$$SSEE = \sum_{i=1}^{m} \sum_{j \neq i=1}^{m} \sum_{k=1}^{n_{ij}} \left(y_{ijk} - \bar{y}_{ij.} \right) \left(y_{jik} - \bar{y}_{ji.} \right).$$

With the 7×1 vector $S = (SSA, SSB, SSAB, SSG, SSGG, SSE, SSEE)'$, we have $\mathrm{E}(S) = C\theta$. The coefficient matrix C is obtained as we derive the expected values (under the round robin model) of each of the sums of squares as linear combinations of the variance-covariance parameters. The ANOVA estimator of θ is then given by $\widehat{\theta} = C^{-1}S$, provided that C is non-singular. Straightforward but lengthy algebra leads us to the matrix C presented in the Appendix. Although useful conditions for the non-singularity of C have not been established, we remark that we encountered no difficulties with various examples including the one presented in Section 5.

Analytical derivation of standard errors of ANOVA estimators for unbalanced designs is practically impossible. A practical solution in such situations is the use of resampling methods such as the jackknife and the bootstrap. Whereas implementing a delete-d jackknife is an easy exercise, for highly nonlinear functions (such as those used to estimate the variance-covariance parameters) the jackknife can be inefficient (Efron & Tibshirani 1993, p. 146). As we are dealing with a parametric model, the use of the parametric bootstrap (Efron & Tibshirani 1993, p. 306) for estimating the sampling distribution and variance of $\widehat{\theta}$ seems a reasonable approach. We assume a normal distribution for the random effects in model (1).

Once the ANOVA estimators $\widehat{\theta}$ and $\hat{\mu} = \bar{y}_{...}$ are obtained, we draw B samples $\{ y_{ijk} : k = 1, \ldots, n_{ij}; 1 \leq i \neq j \leq m\}$ from the normal model according to (1) and (2). For each sample we calculate the ANOVA estimate of θ. This provides a realisation from the approximate sampling distribution of $\widehat{\theta}$. The empirical sampling distribution can then be used for estimating the standard errors of the ANOVA estimators.

3. MAXIMUM LIKELIHOOD ESTIMATION

We outline an approach to maximum likelihood estimation based on the normal distribution which applies to unbalanced and incomplete designs. The method is based on scoring and is straightforward to implement. To apply the standard scoring method, we need to write model (1) in matrix notation. By suitably arranging the response variable in a column vector y, the round robin model may be written as

$$y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = X\mu + \begin{pmatrix} Z_1 & Z_2 \\ Z_2 & Z_1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} Z_3 & 0 \\ 0 & Z_3 \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}.$$

Here $y_1 = (y_{121}, \ldots, y_{12n_{12}}, \ldots, y_{(m-1)m1}, \ldots, y_{(m-1)mn_{m-1m}})'$ and $y_2 = (y_{211}, \ldots, y_{21n_{12}}, \ldots, y_{m(m-1)1}, \ldots, y_{m(m-1)n_{m-1m}})'$; the vectors ε_1 and ε_2 defined analogously. The vector μ contains all the fixed effects in addition to the general mean, and X is the design matrix for the fixed effects. The random effects are $\alpha = (\alpha_1, \ldots, \alpha_m)'$, $\beta = (\beta_1, \ldots, \beta_m)'$, $\gamma_1 = (\gamma_{12}, \ldots, \gamma_{1m}, \ldots, \gamma_{(m-1)1}, \ldots, \gamma_{(m-1)m})'$ and $\gamma_2 = (\gamma_{21}, \ldots, \gamma_{m1}, \ldots, \gamma_{1(m-1)}, \ldots, \gamma_{m(m-1)})'$.

Our assumptions about the model parameters in (2) give the following covariance matrices

$$\operatorname{cov}\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \sigma_\alpha^2 I_m & \sigma_{\alpha\beta} I_m \\ \sigma_{\alpha\beta} I_m & \sigma_\beta^2 I_m \end{pmatrix}, \quad \operatorname{cov}\begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} = \begin{pmatrix} \sigma_\gamma^2 I_m & \sigma_{\gamma\gamma} I_m \\ \sigma_{\gamma\gamma} I_m & \sigma_\gamma^2 I_m \end{pmatrix}$$

and

$$\operatorname{cov}\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} = \begin{pmatrix} \sigma_\varepsilon^2 I_m & \sigma_{\varepsilon\varepsilon} I_m \\ \sigma_{\varepsilon\varepsilon} I_m & \sigma_\varepsilon^2 I_m \end{pmatrix}.$$

After some algebraic manipulations, the covariance matrix of y is seen to be

$$\begin{aligned} V &= \operatorname{cov}(y) = \sigma_\alpha^2 \begin{pmatrix} Z_1 Z_1' & Z_1 Z_2' \\ Z_2 Z_1' & Z_2 Z_2' \end{pmatrix} + \sigma_\beta^2 \begin{pmatrix} Z_2 Z_2' & Z_2 Z_1' \\ Z_1 Z_2' & Z_1 Z_1' \end{pmatrix} \\ & + \sigma_{\alpha\beta} \begin{pmatrix} Z_1 & Z_2 \\ Z_2 & Z_1 \end{pmatrix} \begin{pmatrix} Z_2 & Z_1 \\ Z_1 & Z_2 \end{pmatrix}' + \sigma_\gamma^2 \begin{pmatrix} Z_3 Z_3' & 0 \\ 0 & Z_3 Z_3' \end{pmatrix} \\ & + \sigma_{\gamma\gamma} \begin{pmatrix} 0 & Z_3 Z_3' \\ Z_3 Z_3' & 0 \end{pmatrix} + \sigma_\varepsilon^2 \begin{pmatrix} I_N & 0 \\ 0 & I_N \end{pmatrix} + \sigma_{\varepsilon\varepsilon} \begin{pmatrix} 0 & I_N \\ I_N & 0 \end{pmatrix}. \end{aligned} \tag{3}$$

Under the assumption of normality of the random effects, the log-likelihood of y (ignoring additive constant terms) is

$$\ell(\mu, \sigma_\alpha^2, \sigma_\beta^2, \sigma_\gamma^2, \sigma_\varepsilon^2, \sigma_{\alpha\beta}, \sigma_{\gamma\gamma}, \sigma_{\varepsilon\varepsilon}) = -\frac{1}{2} \log |V| - \frac{1}{2} (y - X\mu)' V^{-1} (y - X\mu)$$

and the MLE for the fixed effects is the generalized least squares (GLS) estimator

$$\hat{\mu} = (X' V^{-1} X)^{-1} X' V^{-1} y.$$

For the variance-covariance parameters $\theta = (\sigma_\alpha^2, \sigma_\beta^2, \sigma_\gamma^2, \sigma_\varepsilon^2, \sigma_{\alpha\beta}, \sigma_{\gamma\gamma}, \sigma_{\varepsilon\varepsilon})'$, let $(\partial V_i / \partial \theta_i)$ be the matrix of elementwise derivatives of the entries of the matrix V with respect to θ_i. As the matrix V in (3) is a linear combination of matrices, the derivative $\partial V_i / \partial \theta_i$ is easy to obtain. For example, the derivative with respect to $\sigma_{\alpha\beta}$ is given by

$$\frac{\partial V}{\partial \sigma_{\alpha\beta}} = \begin{pmatrix} Z_1 & Z_2 \\ Z_2 & Z_1 \end{pmatrix} \begin{pmatrix} Z_2 & Z_1 \\ Z_1 & Z_2 \end{pmatrix}'.$$

For the parameter θ_i, the score equation is given by (Searle *et al.* 1992, p. 384)

$$U_i = \left.\frac{\partial \ell}{\partial \theta_i}\right]_{\mu = \hat{\mu}} = -\frac{1}{2} \operatorname{tr}\left(V^{-1} \frac{\partial V}{\partial \theta_i} \right) + \frac{1}{2} (y - X\hat{\mu})' V^{-1} \frac{\partial V}{\partial \theta_i} V^{-1} (y - X\hat{\mu})$$

and the (i,j)th entry of the information matrix is

$$H_{ij} = -\mathrm{E}\left(\frac{\partial^2 \ell}{\partial\theta_i \partial\theta_j}\right) = \frac{1}{2}\,\mathsf{tr}\left(V^{-1}\frac{\partial V}{\partial\theta_i}V^{-1}\frac{\partial V}{\partial\theta_j}\right).$$

The score equations are solved iteratively. At the $(h+1)th$ iteration step, the update is given by

$$\hat{\theta}^{(h+1)} = \hat{\theta}^{(h)} + \left(\hat{H}^{(h)}\right)^{-1} U^{(h)}$$

with $\hat{\mu}$, the score vector $U = (U_1, \ldots, U_7)'$ and the information matrix H updated at every iteration step.

An estimate of the large sample covariance matrix of the maximum likelihood estimator of θ is $\widehat{H}^{-1}$. The score equations can be computationally intensive as the calculations involve the multiplication and inversion of $2N \times 2N$ matrices. Although the theory and steps given here are straightforward, they do not appear to have been previously recorded in the round robin literature.

3.1 Empirical Bayes estimation of random effects

The estimates of individual random effects may be of interest. For example, when ranking team i relative to team j in a sports competition, one would be interested in $(\alpha_i, \beta_j, \gamma_{ij})$ and $(\alpha_j, \beta_i, \gamma_{ji})$. For the sake of simplicity, let us consider the round robin model without the interaction effects $\gamma_{ij}, i, j = 1, \ldots, m$. The general model can be dealt with similarly. We have

$$y = X\mu + \begin{pmatrix} Z_1 & Z_2 \\ Z_2 & Z_1 \end{pmatrix}\begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

with

$$D = \mathrm{cov}\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \sigma_\alpha^2 I_m & \sigma_{\alpha\beta} I_m \\ \sigma_{\alpha\beta} I_m & \sigma_\beta^2 I_m \end{pmatrix} \quad \text{and} \quad \Omega = \mathrm{cov}\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} = \begin{pmatrix} \sigma_\varepsilon^2 I_m & \sigma_{\varepsilon\varepsilon} I_m \\ \sigma_{\varepsilon\varepsilon} I_m & \sigma_\varepsilon^2 I_m \end{pmatrix}$$

so that

$$V = \mathrm{cov}(y) = \begin{pmatrix} Z_1 & Z_2 \\ Z_2 & Z_1 \end{pmatrix} D \begin{pmatrix} Z_1 & Z_2 \\ Z_2 & Z_1 \end{pmatrix}' + \Omega = ZDZ' + \Omega.$$

An empirical Bayes estimator of the actor and partner effects given by the posterior mean is then (Speed 1991)

$$\begin{pmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{pmatrix} = \hat{D}Z'\hat{V}^{-1}(y - X\hat{\mu}),$$

where $\hat{D}$, $\hat{V}$ and $\hat{\mu} = (X'V^{-1}X)^{-1}X'V^{-1}y$ are estimated using the maximum likelihood algorithm and

$$\widehat{\mathrm{cov}}\begin{pmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{pmatrix} = \hat{D} - \hat{D}Z'\hat{V}^{-1}Z\hat{D} + \hat{D}Z'\hat{V}^{-1}X(X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}Z\hat{D}.$$

4. BAYESIAN ANALYSIS

In this section, we outline a fully Bayesian approach to the analysis of round robin interaction data. A fully Bayesian analysis has conceptual appeal in that priors are constructed from an understanding of the structure of the experiment, from subjective opinions and from invariance considerations. In contrast, empirical Bayes procedures as utilized by Wong (1982) and studied in Section 3 employ priors that are based on observed data. Although there are often few alternatives to empirical Bayes procedures, the reliance of priors on data is in fact in conflict with the Bayesian paradigm.

We will see that our fully Bayesian approach also has practical advantages. For example, in contrast to the methods of Section 3, we no longer face computational difficulties associated with the inversion of large matrices as our approach relies only on conditional distributions having fixed dimension. This is a by-product of using the Gibbs sampling algorithm. Also, the Bayesian analysis is based on the exact posterior distribution rather than an asymptotic normal distribution in the case of maximum likelihood estimation. In addition, the fully Bayesian approach can handle unbalanced and incomplete designs with no added difficulty. The methodology presented here may be generalized in certain directions by introducing additional covariates and their associated prior distributions. For example, one may wish to introduce a subscript k on the mean μ to denote a temporal trend. Finally, our fully Bayesian approach based on simulation from the posterior permits the investigation of any posterior characteristic of interest including marginal posterior distributions. This is in contrast with the classical analyses of Sections 2 and 3 which focus only on the estimation of primary parameters and their standard errors.

Our basic model is the traditional round robin model (1) with additional distributional assumptions placed on the data, parameters and hyperparameters. We let $\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$ and following Wong (1982), we assume conditionally

$$\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} \sim \text{Normal}_2 \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix},\ \Sigma_1 \right],$$

$$\begin{pmatrix} \gamma_{ij} \\ \gamma_{ji} \end{pmatrix} \sim \text{Normal}_2 \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix},\ \Sigma_2 \right],$$

$$\begin{pmatrix} y_{ijk} \\ y_{jik} \end{pmatrix} \sim \text{Normal}_2 \left[\begin{pmatrix} \mu_{ij} \\ \mu_{ji} \end{pmatrix},\ \Sigma_3 \right]$$

where $k = 1, \ldots, n_{ij}$, $1 \le i \ne j \le m$ and

$$\Sigma_1 = \begin{pmatrix} \sigma^2_\alpha & \sigma_{\alpha\beta} \\ \sigma_{\alpha\beta} & \sigma^2_\beta \end{pmatrix},\quad \Sigma_2 = \begin{pmatrix} \sigma^2_\gamma & \sigma_{\gamma\gamma} \\ \sigma_{\gamma\gamma} & \sigma^2_\gamma \end{pmatrix},\quad \Sigma_3 = \begin{pmatrix} \sigma^2_\varepsilon & \sigma_{\varepsilon\varepsilon} \\ \sigma_{\varepsilon\varepsilon} & \sigma^2_\varepsilon \end{pmatrix}.$$

It is at this point where our modelling assumptions differ from Wong (1982) and take on an extended hierarchical structure. Following conventional Bayesian protocol for linear models (Gelfand, Hills, Racine-Poon & Smith 1990), we assume

$$\begin{aligned} \mu &\sim \text{Normal}[\tau, \sigma^2_\mu],\quad \Sigma_1^{-1} \sim \text{Wishart}_2[(\rho_0 R)^{-1}, \rho_0], \\ \tau &\sim \text{Normal}[\tau_0, \sigma^2_\tau],\quad \sigma^2_\mu \sim \text{Inverse Gamma}[a_0, b_0], \end{aligned}$$

where the hyperparameters are set to give diffuse prior distributions for the parameters τ, σ^2_μ and Σ_1. For example, as described in Section 5, we set $\tau_0 = 0$, $\sigma_\tau = 10000$, $a_0 = 0.0001$, $b_0 = 0.0001$, $\rho_0 = 2$ and $R = r_0 I$ where r_0 is determined by the experimental structure.

The prior assumptions on the covariance matrices Σ_2 and Σ_3 are atypical due to the necessity of equal diagonal entries. In fact, this is one of the ways in which the assumptions for the round robin model render its analysis nonstandard for the popular Bayesian software package BUGS (Spiegelhalter, Thomas, Best & Gilks 1996). For Σ_2, we assume that $\sigma^2_\gamma \sim \text{Exponential}[r_0]$ and $\sigma_{\gamma\gamma} \mid \sigma_\gamma \sim \text{Uniform}[-\sigma^2_\gamma, \sigma^2_\gamma]$. The conditional uniform prior is vague and is motivated by the restriction $-1 \le \rho_2 \le 1$. A similar prior structure is also imposed on Σ_3. These atypical bivariate priors are appealing as they are simply characterized by the single specified hyperparameter r_0.

Moreover, $\mathrm{E}(\sigma_\alpha^2) = \mathrm{E}(\sigma_\beta^2) \approx \mathrm{E}(\sigma_\varepsilon^2) = \mathrm{E}(\sigma_\gamma^2) = r_0$ which suggests a commonality of magnitude amongst the various effects. Finally, the experimental structure usually dictates that we truncate μ_{ij} to some interval (k_{10}, k_{20}). For example, with nonnegative data y_{ijk} we would impose the prior restriction $(k_{10}, k_{20}) = (0, \infty)$.

Letting $[A \mid B]$ denote the conditional distribution of A given B, it follows from our modelling assumptions that the posterior distribution

$$\begin{aligned}[\mu, \alpha, \beta, \gamma, \theta, \tau, \sigma_\mu^2 \mid y] \quad \propto \quad & [y \mid \mu, \alpha, \beta, \gamma, \sigma_\varepsilon^2, \sigma_{\varepsilon\varepsilon}]\ [\mu \mid \tau, \sigma_\mu^2]\ [\alpha, \beta \mid \sigma_\alpha^2, \sigma_\beta^2, \sigma_{\alpha\beta}] \\ & [\gamma|\sigma_\gamma^2, \sigma_{\gamma\gamma}]\ [\sigma_\varepsilon^2, \sigma_{\varepsilon\varepsilon}]\ [\tau]\ [\sigma_\mu^2]\ [\sigma_\alpha^2, \sigma_\beta^2, \sigma_{\alpha\beta}]\ [\sigma_\gamma^2, \sigma_{\gamma\gamma}], \end{aligned} \tag{4}$$

where $\theta = (\sigma_\alpha^2, \sigma_\beta^2, \sigma_\gamma^2, \sigma_\varepsilon^2, \sigma_{\alpha\beta}, \sigma_{\gamma\gamma}, \sigma_{\varepsilon\varepsilon})'$ is the vector of variance-covariance parameters of primary interest as previously defined. We therefore have a challenging posterior distribution of dimension $(m^2 + m + 10)$ where α and β are m-dimensional and γ is $(m^2 - m)$-dimensional. The model is driven by $N = \sum_{1 \le i < j \le m} n_{ij}$ observations.

Our goal now is to simulate from the posterior distribution (4) so that marginal posterior characteristics can be estimated. To do so, we use the Gibbs sampling algorithm, which is an iterative approach to simulation from a target distribution. Gibbs sampling has been successfully used in many Bayesian problems involving high dimensionality (Gelfand & Smith 1990). Based on the posterior structure (4), an implementation of Gibbs sampling proceeds by generating from the full conditional distributions

$$\begin{gathered}[\mu \mid \cdot\,],\quad [\alpha_i, \beta_i \mid \cdot\,],\quad [\gamma_{ij}, \gamma_{ji} \mid \cdot\,],\quad [\tau \mid \cdot\,],\quad [\sigma_\mu^2 \mid \cdot\,], \\ [\sigma_\alpha^2, \sigma_\beta^2, \sigma_{\alpha\beta} \mid \cdot\,],\quad [\sigma_\varepsilon^2, \sigma_{\varepsilon\varepsilon} \mid \cdot\,],\quad [\sigma_\gamma^2, \sigma_{\gamma\gamma} \mid \cdot\,], \end{gathered}$$

where $1 \le i < j \le m$ and $[A \mid \cdot\,]$ denotes the conditional distribution of A given all other parameters and the data y. The first six conditional distributions are respectively Normal, Normal$_2$, Normal$_2$, Normal, Inverse Gamma and Wishart$_2$. Although tedious, the derivation of the parameters for these distributions is straightforward and is available from the authors upon request. The remaining conditional distributions $[\sigma_\varepsilon^2, \sigma_{\varepsilon\varepsilon} \mid \cdot\,]$ and $[\sigma_\gamma^2, \sigma_{\gamma\gamma} \mid \cdot\,]$ have the same nonstandard form and we outline the simulation algorithm for $[\sigma_\varepsilon^2, \sigma_{\varepsilon\varepsilon} \mid \cdot\,]$. From (4), we note that the density of $[\sigma_\varepsilon^2, \sigma_{\varepsilon\varepsilon} \mid \cdot\,]$ is proportional to

$$f(\Sigma_3)\ I_{(-\sigma_\varepsilon^2, \sigma_\varepsilon^2)}(\sigma_{\varepsilon\varepsilon})\ \frac{1}{r_0} \exp(-\sigma_\varepsilon^2 / r_0),$$

where $I_{(a,b)}$ is the indicator function on the interval (a, b) and

$$f(\Sigma_3) = |\Sigma_3|^{-N/2} \exp\left\{ -\frac{1}{2} \sum_{i<j} \sum_{k=1}^{n_{ij}} \begin{pmatrix} y_{ijk} - \mu_{ij} \\ y_{jik} - \mu_{ji} \end{pmatrix}' \Sigma_3^{-1} \begin{pmatrix} y_{ijk} - \mu_{ij} \\ y_{jik} - \mu_{ji} \end{pmatrix} \right\}.$$

Therefore, within the Gibbs sampling algorithm we can "embed" a Metropolis step whereby we generate $u \sim$ Uniform$[0, 1]$ and new variates $\sigma_\varepsilon^2 \sim$ Exponential$[r_0]$ and $\sigma_{\varepsilon\varepsilon} \sim$ Uniform$[-\sigma_\varepsilon^2, \sigma_\varepsilon^2]$. We stick on the "old" value of Σ_3 if $u > f(\Sigma_3^{\text{new}}) / f(\Sigma_3^{\text{old}})$. Although this independence sampler may have low acceptance rates, it is simple to code and is adequate for the application considered in this paper and others studied by the authors.

As a final note, we remark that the fully Bayesian methodology presented in this section permits the simple investigation of sub-models. For if the experimenter determines that a certain parameter is unimportant, then the parameter can be eliminated from the analysis by simply substituting constant null values when it would ordinarily be generated from its full conditional distribution.

5. SOCIAL INTERACTION DATA EXAMPLE

In this section, we discuss the analysis of a balanced round robin experiment. Warner (1978) conducted a round robin study involving eight subjects. Each pair of subjects conversed privately on three separate occasions for about 12 to 15 minutes and the percent of time spent speaking by each subject was the response variable. The details of the experimental setup and the raw data are given in Warner *et al.* (1979). Due to recording/measurement errors and both persons talking or remaining silent simultaneously, the percents do not typically add to 100. In fact, the sums are often as small as 75, and as large as 125.

Following Wong (1982), we provide a short description of the parameters of the round robin model in this setting. The actor effect α_i represents person i's talkativeness, and the partner effect β_i measures that person's ability to elicit conversation. The interaction effect γ_{ij} represents the special adjustment that person i makes in level of talkativeness when paired with person j. In this experiment, the subjects were not well acquainted with one another. Therefore, we do not expect γ_{ij} to dominate the main effects α_i and β_j. The variance components σ^2_α and σ^2_β measure the variability in the talkativeness and listening capability, respectively. The parameter $\sigma_{\alpha\beta}$ measures the covariance between a person's speech activity level and the effect on the partner's activity level. Naturally, we would expect $\sigma_{\alpha\beta}$ to be negative as excessive talking precludes listening and vice-versa. The parameter σ^2_γ measures the variability in special adjustments amongst the pairs of subjects. The covariance $\sigma_{\gamma\gamma}$ represents the degree to which the conversation is stimulating; when $\sigma_{\gamma\gamma} > 0$, this indicates that both parties are anxious to speak and there is little deadtime. The error variance σ^2_ε measures the contribution of measurement errors and other situational factors. The covariance $\sigma_{\varepsilon\varepsilon}$ is expected to be negative as the data are recorded in percents. We note that the above parameters may have completely different interpretations depending on the application (e.g., the analysis of scores from sporting contests).

Table 1 shows the estimates of covariance components and their standard errors as obtained from ANOVA based estimation and maximum likelihood estimation. The standard errors of ANOVA estimates are based on parametric bootstrap sampling as described in Section 2. The standard errors of the maximum likelihood estimates are based on the estimated large sample covariance matrix. Except for σ^2_α, the estimates and the standard errors are very similar. Except for the error parameters σ^2_ε and $\sigma_{\varepsilon\varepsilon}$, we observe that none of the parameters are strongly significant, and this might be expected with such a small sample. We also note that σ^2_α is twice σ^2_β indicating that there is more variation in talkativeness than in the ability to elicit conversation. The correlation between α and β is $-.66$ (ANOVA) and $-.71$ (MLE) which confirms the high degree to which talking limits one's listening.

Table 1: Parameter estimates and standard errors.

Parameter	ANOVA		MLE		Bayes ($r_0 = 70$)		Bayes ($r_0 = 40$)	
	Est	SE	Est	SE	Est	SD	Est	SD
σ^2_α	92.0	56.4	84.2	49.3	129.4	98.1	116.2	91.5
σ^2_β	40.9	28.5	40.7	30.2	76.2	56.8	62.3	48.5
$\sigma_{\alpha\beta}$	-40.4	33.8	-41.6	32.5	-42.3	57.6	-42.4	51.3
σ^2_γ	30.0	19.7	30.1	22.5	35.4	18.6	33.7	16.9
$\sigma_{\gamma\gamma}$	4.1	19.8	2.9	22.5	0.0	17.2	1.2	15.4
σ^2_ε	146.1	23.5	146.0	22.6	146.2	21.2	142.2	19.7
$\sigma_{\varepsilon\varepsilon}$	-95.5	23.1	-95.5	22.6	-91.9	21.2	-88.5	19.4

We also record in Table 1 estimates of the posterior means and posterior standard deviations from a fully Bayesian analysis. As discussed in Section 4, the prior distribution is completely determined by specifying a value for the hyperparameter r_0. We argue that μ_{ij} should be centred roughly at $\mu = 50$ with α_i, β_j and γ_{ij} combining to give a maximum effect of 50. Since α_i and β_j contribute in the opposite direction for the maximum effect, we expect a maximum absolute effect of 25 for either α_i or β_j. In other words, $3\sigma_\alpha \approx 25$, and since $\mathrm{E}(\sigma_\alpha^2) \approx r_0$, we set $r_0 = (25/3)^2 = 70$. Note that this gives a fully Bayesian analysis as we have used the structure of the experiment to determine the prior value for r_0.

We observe that the fully Bayesian estimates are in the same general direction as those obtained via the ANOVA and maximum likelihood procedures. We mention that if we change r_0 considerably (e.g., $r_0 = 40$, see Table 1), similar results are obtained except for the parameters σ_α^2 and σ_β^2. Borrowing ideas from the theory of linear models, we have only $m-1$ degrees of freedom for each of σ_α^2 and σ_β^2, and thus, we should not be surprised that the data do not overwhelm the priors. On the other hand, σ_ε^2 and σ_γ^2 have more degrees of freedom and are not very sensitive to the choice of r_0. An alternative way of specifying r_0 for general round robin applications, is to use an empirical Bayes approach based on ANOVA or maximum likelihood estimates such as $\hat{\sigma}_\alpha$ or $\hat{\sigma}_\beta$.

As a practical concern, we observed strong autocorrelations between successive updates in the Gibbs sampling algorithm. To counteract this, the generated output was thinned by choosing every 1000th variate. This was particularly needed for the parameters σ_ε^2, $\sigma_{\varepsilon\varepsilon}$, σ_γ^2 and $\sigma_{\gamma\gamma}$ for which the Metropolis steps proved to be sticky. The results reported in Table 1 are based on 5000 stored variates and this required 45 minutes of computation on a Sun workstation.

In Table 2, we present the empirical Bayes estimates for the actor and partner effects as described in Section 3.1. We also include both sets of the fully Bayes estimates and observe close agreement amongst all three analyses. It therefore seems that there is less sensitivity amongst the "first-order" parameters than the "second-order" variance-covariance parameters from Table 1.

Table 2: Empirical Bayes and fully Bayes estimates of the actor and partner effects.

Subject	Empirical Bayes				Bayes ($r_0 = 70$)				Bayes ($r_0 = 40$)			
i	$\widehat{\alpha}_i$	SE	$\widehat{\beta}_i$	SE	$\widehat{\alpha}_i$	SD	$\widehat{\beta}_i$	SD	$\widehat{\alpha}_i$	SD	$\widehat{\beta}_i$	SD
1	13.8	4.0	-5.7	3.1	13.3	5.1	-5.3	4.2	13.0	4.9	-5.3	4.0
2	-12.3	4.0	1.5	3.1	-11.3	5.1	1.3	4.2	-11.2	4.9	1.5	4.0
3	-13.1	4.0	10.6	3.1	-12.4	5.1	10.2	4.4	-12.4	4.9	9.9	4.1
4	2.9	4.0	-6.9	3.1	3.2	5.0	-5.8	4.2	3.2	4.8	-5.5	3.9
5	8.1	4.0	-5.5	3.1	7.9	5.1	-5.2	4.3	7.9	4.9	-5.1	4.0
6	-3.7	4.0	-3.5	3.1	-3.3	5.1	-3.4	4.2	-3.1	4.8	-3.2	3.9
7	-0.2	4.0	2.7	3.1	-0.2	5.0	2.8	4.2	-0.2	4.8	2.6	3.9
8	4.5	4.0	5.9	3.1	4.1	5.0	5.9	4.3	3.9	4.8	5.5	4.0

APPENDIX

Let $C = (c_{ij})$ be the matrix for ANOVA estimation discussed in Section 2. We have

$c_{11} = c_{22} = c_{41} = c_{42} = 2N - \sum n_{i.}^2/(2N)$,
$c_{12} = \sum\sum n_{ij}^2/n_{i.} - \sum n_{i.}^2/(2N)$,
$c_{13} = c_{23} = c_{43} = -\sum n_{i.}^2/N$,
$c_{14} = c_{35} = \sum\sum n_{ij}^2\{1/n_{i.} - 1/(2N)\}$,
$c_{15} = c_{25} = c_{34} = c_{45} = c_{54} = -\sum\sum n_{ij}^2/(2N)$,
$c_{16} = c_{26} = c_{37} = m - 1$,
$c_{17} = c_{27} = c_{36} = c_{47} = c_{56} = -1$,

$c_{21} = \sum\sum n_{ij}^2/n_{.j} - \sum n_{i.}^2/(2N)$,
$c_{24} = \sum\sum n_{ij}^2\{1/n_{.j} - 1/(2N)\}$,
$c_{31} = c_{32} = c_{51} = c_{52} = -\sum n_{i.}^2/(2N)$,
$c_{33} = 2N + \sum\sum n_{ij}^2/n_{i.} - \sum n_{i.}^2/N$,
$c_{44} = c_{55} = 2N - \sum\sum n_{ij}^2/(2N)$,
$c_{46} = c_{57} = m(m-1) - 1$,
$c_{53} = 4N - \sum n_{i.}^2/N$,
$c_{61} = c_{62} = c_{63} = c_{64} = c_{65} = c_{67} = c_{71} = c_{72} = c_{73} = c_{74} = c_{75} = c_{76} = 0$,
$c_{66} = c_{77} = 2N - m(m-1)$.

ACKNOWLEDGEMENTS

The authors thank the Editor and two referees for their comments and encouragement. P. Gill thanks the Okanagan University College for a sabbatical leave during which the work was initiated. Both authors were supported by grants from the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

C. F. Bond & B. R. Lashley (1996). Round-robin analysis of social interaction: Exact and estimated standard errors. *Psychometrika*, 61, 303–311.

B. Efron & R. Tibshirani (1993). *An Introduction to the Bootstrap*. Chapman & Hall, New York.

A. E. Gelfand & A. F. M. Smith (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85, 398–409.

A. E. Gelfand, S. E. Hills, A. Racine-Poon & A. F. M. Smith (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association*, 85, 972–985.

D. A. Kenny (1994). *Interpersonal Perception: A Social Relations Analysis*. Guilford Press, New York.

B. R. Lashley & C. F. Bond (1997). Significance testing for round robin data. *Psychological Methods*, 2, 278–291.

J. Lev & E. Kinder (1957). New analysis of variance formulas for treating data from mutually paired subjects. *Psychometrika*, 22, 1–15.

S. R. Searle, G. Casella & C. McCulloch (1992). *Variance Components*. Wiley, New York.

T. Speed (1991). Discussion of G. K. Robinson, 1991. *Statistical Science*, 6, 42–44.

D. J. Spiegelhalter, A. Thomas, N. Best & W. R. Gilks (1996). BUGS: Bayesian Inference Using Gibbs Sampling, Version 0.5 (version ii).

S. M. Stigler (1994). Citation patterns in the journals of statistics and probability. *Statistical Science*, 9, 94–108.

R. Warner (1978). *Temporal Patterns in Dialogue*. Unpublished doctoral dissertation, Harvard University, Boston, MA.

R. M. Warner, D. A. Kenny & M. Stoto (1979). A new round robin analysis of variance for social interaction data. *Journal of Personality & Social Psychology*, 37, 1742–1757.

S. Wasserman & K. Faust (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge, UK.

G. Y. Wong (1982). Round robin analysis of variance via maximum likelihood. *Journal of the American Statistical Association*, 77, 714–724.

Received 9 November 2000
Accepted 26 January 2001

Paramjit S. GILL: pgill@ouc.bc.ca
Department of Mathematics and Statistics
Okanagan University College
Kelowna, British Columbia
Canada V1V 1V7

Tim B. SWARTZ: tim@stat.sfu.ca
Department of Statistics and Actuarial Science
Simon Fraser University
Burnaby, British Columbia
Canada V5A 1S6